HAMS Data Science Challenge

Maciej Combrzynski-Nogala

January 2023

In this report, I answer questions for HAMS Data Science Challenge. The challenge aims to apply the Bayesian mixed-media model (MMM) to the test data and interpret the results. MMMs are used by advertisers to measure the effectiveness of their advertising campaigns and provide insight into making future budget allocation decisions. For this specific use case, the model is estimated using the Bayesian approach.

1 How do you model spend carryover?

In advertising, it is believed that it takes some time until the advertisement starts affecting the target audience. The carryover effect is the lag after which the advertisement spending is effective. In my solution, I am modeling it using *adstock* function:

$$\text{adstock}\left(x_{t-L+1,m},\ldots,x_{t,m};w_m,L\right)=\frac{\sum_{l=0}^{L-1}w_m(l)x_{t-l,m}}{\sum_{l=0}^{L-1}w_m(l)}, \tag{1}$$

where w_m is a nonnegative weight function; *L* is the maximum duration of the carryover effect; $x_{t,m}$ it the media spend of channel *m* at week *t*. Weight function can have different forms, the most common ones are *geometric decay* and *delayed adstock*. Geometric decay assumes that from the moment of the spend, its effect will be decreasing geometrically. It is calculated using the following formula:

$$w_m^g\left(l;\alpha_m\right)=\alpha_m^l,\quad l=0,\ldots,L-1,\quad 0<\alpha_m<1, \tag{2}$$

where α is the retention rate.

Delayed adstock assumes that for a certain period of time, the effect will be increasing, and after reaching the peak, it will start decreasing. It is calculated using the formula:

$$w_m^d\left(l,\alpha_m,\theta_m\right)=\alpha_m^{(l-\theta_m)^2},\quad l=0,\quad L-1.\quad 0<\alpha_m<1\quad 0\le\theta_m\le L-1, \tag{3}$$

where α is the retention rate and θ is the delay peak effect.

For modelling I am using $L=13$, since with larger values, weights converge to 0, and its effect is very limited anyway.

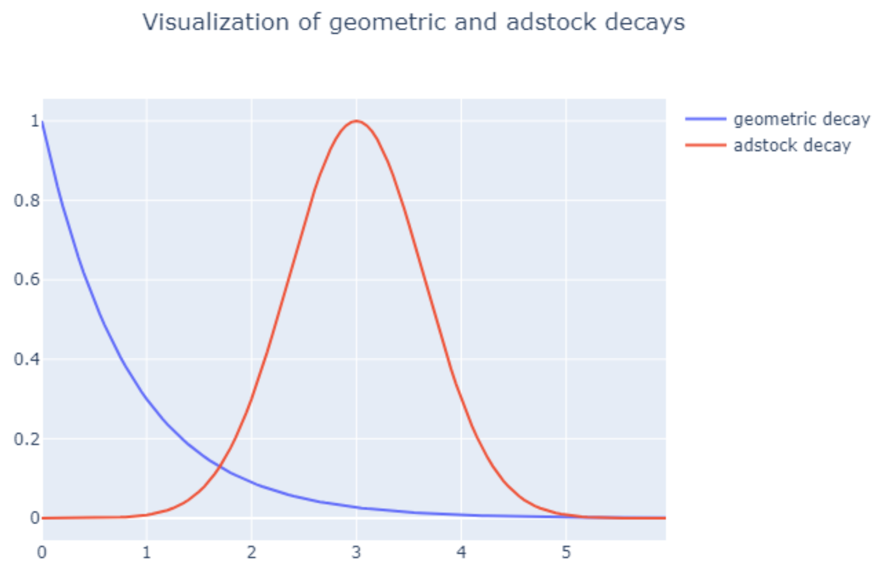

Figure 1: Diagram of geometric and delayed adstock decays. Geometric decay has parameter $\alpha = 0.3$, which can be observed at timestep 1, where the value of decay is equal to 0.3. Delayed adstock is modeled with the same alpha and $\theta = 3$, which defines the peak moment.

2 Explain your choice of prior inputs to the model?

To model the revenue, on top of the adstock transformed spend data, I am also using control variables, trends, and seasonality of the spend. I decomposed each time series using STL decomposition and then added it to the model. I based the model structure on the paper *Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects* by Jin et al. The model I am using takes the following form:

$$y_t = \sum_{m=1}^{M} \beta_m \operatorname{adstock}(x_{t,m}) + \sum_{c=1}^{C} \gamma_c z_{t,c} + \epsilon_t, \tag{4}$$

where γ_c is the effect of the control variable z_c and ϵ_t is the noise term. Normally to get prior distributions I would contact someone with industry experience or use previous models, but in this case, I do not have access to either of them. The following table shows the distribution types for each of them:

Parameter	Prior	Parameter	Prior
α	beta(3,3)	γ	normal $(0, 1)$
θ	uniform $(0, 12)$	Var(ϵ)	inverse gamma $(0.05, 0.0005)$
β	half normal $(0, 1)$		

Table 3: Priors on parameters.

I am using the priors from the above mention paper, and I will try to explain the reasoning behind them:

- α - beta(3,3) is constrained between [0, 1], with alpha=3, and beta=3 it peaks at 0.5. It is a good choice if we do not know exactly how fast the decay occurs.

- θ - uniform(0,12) allows us to uniformly sample theta between 0 and L, this lets us see the impact of the lag on the model results. Larger values make weights to be very small, close to 0 which limits their impact on the model.
- - half normal (0,1), β defines the impact of each media on the revenue. I assume that the impact of each media is positive. It is also easy to sample from which is useful when there are many different media.
- γ - normal(0,1), γ defines impact of the control variables. Choice of the normal distribution is the same as with β, however here it is possible that certain seasons or trends have a negative impact on the revenue, thus its normal distribution that allows negative values.
- $\mathrm{Var}(\epsilon)$ - inverse gamma (0.05,0.0005), is a popular way to model the error term for Bayesian methods.

3 How are your model results based on prior sampling vs. posterior sampling?

Unfortunately, I was traveling for the past week and was not able to generate model results on time. Sampling on my machine takes really long time, thus I am not able to answer the rest of the questions based on the data but when possible I will try to briefly explain the steps I would take to answer them.

I would make two predictions, one using parameters obtained using prior distributions and another one using posterior distributions. I would then plot them together with the actual revenue and compare the results.

4 How good is your model performing? How do you measure it?

I would split the original data into a test and train set. Then train the model just on the train set and then see how it performs on the test set alone. One of the problems of MMM is that the amount of data is very limited, to make models more robust it is a good idea to simulate synthetic data and use it to train and test the model.

5 Can you derive ROI (return on investment) estimates per channel? What is the best channel in terms of ROI?

To estimate ROI per channel, I would choose periods where I set the spending of a given channel to 0. Then I would calculate ROI by dividing the difference between the model prediction with normal spend values and without, by the actual spend in a given period.